The Locale of Random Sequences

Alex Simpson

LFCS, School of Informatics
University of Edinburgh, UK

The Application Problem in probability

What is the empirical meaning of event E having probability p?

Standard explanation, cf. Kolmogorov (1933):

- Perform large number of repeated trials.
- Proportion of occurrences of event E is practically certain to be close to p.

Practical certainty means with probability very close to 1.
(Of course, can calculate exact probability that observed value lies within ϵ of p.)

What is the empirical meaning of practical certainty???
A vicious circle!

Random sequences

Explaining the empirical meaning of probability amounts to explaining the characteristic properties of sequences of stochastic events.

So need a theory of random sequences.

Intuitively, a sequence $\alpha \in \mathbf{2}^\omega$ (where $\mathbf{2} := \{0, 1\}$) is random if it could potentially arise by tossing a fair coin (we use 0 for heads and 1 for tails) *ad infinitum*.

We seek a mathematical theory of random sequences $\mathbf{R} \subseteq \mathbf{2}^\omega$.

Task is also interesting for more general sequences of stochastic events. E.g., tossing a sequence of biased coins.

Questions such a theory should answer

- Is the sequence 0^ω random?
- Is the bit-sequence of Chaitin's halting probability Ω random?
- Under what conditions does a transducer on infinite sequences preserve randomness?
- Suppose we toss a sequence of biased coins, where the i-th coin has probability $\frac{1}{2} + \delta_i$ of coming up 0 and $\frac{1}{2} - \delta_i$ of coming up 1. Under what conditions on (δ_i) are we guaranteed to obtain a fair random sequence? (van Lambalgen's brainteaser, 1987)

Such questions seem meaningless in probability theory à la Kolmogorov

von Mises' theory of probability

Richard von Mises (1919) developed a mathematical theory of probability, based directly on random sequences.

He postulated that random sequences should be characterised by principles encapsulating two contrasting aspects of a random $\alpha \in \mathbf{2}^\omega$:

- Local irregularity: e.g., cannot predict α_n from $\alpha_0 \ldots \alpha_{n-1}$
- Global regularity: frequencies converge

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \alpha_i \quad = \quad \frac{1}{2}$$

(this embodies the law of large numbers)

In von Mises' theory probability is defined as the limit of observed frequencies.

Difficulties

von Mises' attempt at formalizing random sequences (whence probability) met with mathematical difficulties.

- Local irregularity is formalized as invariance under "admissible place selections". Formulating this precisely is not straightforward.
- The approach is limited in power; e.g., cannot derive the full law of the iterated logarithm from von Mises' principles.
- The theory is a mathematical outcast. Care is needed to avoid inconsistency. It does not sit easily within set theory.

There are also philosophical issues. For example, the law of large numbers is an idealistic rather than empirical assumption, since we can never observe that it holds.

Observable properties

We seek a theory of random sequences based only on our empirical experience of sequences of stochastic events.

At any time, we only see a finite prefix of a random sequence (and this might be *any* finite sequence of digits).

A subset $U \subseteq \mathbf{2}^\omega$ is open (in the Cantor/product topology) if it satisfies:

$$\alpha \in U \implies \exists n. \forall \beta \in \mathbf{2}^\omega. (\beta\lceil_n = \alpha\lceil_n \implies \beta \in U)$$

Open subsets correspond to observable properties, cf. Smyth, Abramsky.

For example, $\{\alpha \mid \alpha \neq 0^\omega\}$ is observable. But $\{0^\omega\}$ is not observable.

Empirical properties of randomness

An empirical fact: If α is a random sequence, then we will sooner or later observe that $\alpha \neq 0^\omega$, i.e., that $\alpha \in \{\beta \in \mathbf{2}^\omega \mid \beta \neq 0^\omega\}$.

Note that, for any n, the probability that $\alpha\lceil_n = 0^n$ is 2^{-n}. Thus the probability that the prefix $\alpha\lceil_n$ fails to show that $\alpha \neq 0^\omega$ is 2^{-n}.

We soon have to have been very unlucky not to have observed that $\alpha \neq 0^\omega$.

The empirical fact is explained by a general principle that the observer is not infinitely unlucky.

We turn this principle into a (first) postulate of randomness.

Postulates of randomness

We postulate properties of the desired set $\mathbf{R} \subseteq \mathbf{2}^\omega$ of random sequences. Let λ be the uniform/Lebesgue measure on $\mathbf{2}^\omega$.

(R1) First postulate of randomness

For every open $U \subseteq \mathbf{2}^\omega$, if $\lambda(U) = 1$ then $\mathbf{R} \subseteq U$.

Informally: if α is a random sequence, and U is an almost sure observable property then we will (eventually) observe that $\alpha \in U$.

The first postulate addresses the Application Problem. The theory predicts that any almost sure observation will actually occur.

Thus "almost sure" becomes "sure" for open sets.

The second postulate is a nontriviality condition. Its purpose is to ensure that there are enough random sequences.

(R2) Second postulate of randomness

For every open $U \subseteq \mathbf{2}^\omega$, if $\mathbf{R} \subseteq U$ then $\lambda(U) = 1$.

Informally: if U is an observable property with $\lambda(U) < 1$ then there has to be a random sequence outside U — since there is some finite probability of randomly generating such a sequence.

(There is some similarity with Myhill's set-theoretic axioms for randomness, cf. van Lambalgen (1992). Myhill's axioms, however, use measurable sets rather than open sets and invoke set-theoretic definability. I believe it is easier to justify the above postulates as embodying empirical properties of randomness.)

Inconsistency

The two postulates of randomness are inconsistent.

For any sequence β, the open set $\{\alpha \mid \alpha \neq \beta\}$ has measure 1.
Therefore, by the first postulate of randomness:

$$\mathbf{R} \subseteq \{\alpha \mid \alpha \neq \beta\} \ ,$$

i.e., $\beta \notin \mathbf{R}$.

Thus $\mathbf{R} = \emptyset$, but this contradicts the second postulate.

Conventional wisdom

Ideally, a random sequence would be one satisfying all measure 1 properties (irrespective of whether open or not).

There are no such sequences.

It is therefore necessary to restrict to a a countable family $\mathcal{F}$ of (not necessarily open) measure 1 subsets of $\mathbf{2}^\omega$.

Define:

$$\mathbf{R}_{\mathcal{F}} := \bigcap \mathcal{F} \ ,$$

Because $\mathcal{F}$ is countable, we have $\lambda(\mathbf{R}_{\mathcal{F}}) = 1$, hence nontriviality.

N.B., our first postulate of randomness is weakened, but the second holds as stated.

Algorithmic and logical notions of randomness

Typically the family $\mathcal{F}$ is given using either recursion theoretic or logical notions of definability. Numerous such definitions have been proposed and argued for; for example, by Martin-Löf (1966), Schnorr (1971), Kurtz (1981), ...

Algorithmic randomness is an important area of recursion theory. Does it, however, provide the right framework for modelling the stochastic phenomenon of randomness?

Two criticisms:

- No one canonical notion of randomness
- While one can make a (dubious!) case that recursion-theoretic restrictions reflect our observational limitations, there is no reason to believe any recursion-theoretic dependencies to be inherent in the stochastic phenomenon of randomness itself

Aim of talk

The goal of the talk is to present a mathematically natural approach to modelling the stochastic phenomenon of randomness.

A notable feature is that our two postulates of randomness will hold as originally formulated.

In particular, there is no need to select a countable family $\mathcal{F}$ of measure 1 sets.

Conventional wisdom is wrong

Some clues

The two postulates of randomness are couched in terms of observable properties (open sets).

Thus it is natural to define randomness in a setting in which observable properties play a fundamental role.

Further, we never experience a completed infinite random sequence in its entirety — only its finite prefixes.

Thus it is natural to define randomness in a setting in which completed entities (points) are not the basic ingredient.

All this suggests using locale theory (a.k.a. point-free topology)

Locales

A locale X is given by a frame $\mathcal{O}(X)$, that is, $\mathcal{O}(X)$ is a partially-ordered set with:

- arbitrary joins $\bigcup$ (including the empty join $\emptyset$),
- (hence) finite meets $\cap$ (including the empty meet X),
- satisfying the distributive law:

$$U \cap (\bigcup_{i \in I} V_i) = \bigcup_{i \in I} U \cap V_i \ .$$

Motivating example: $\mathcal{O}(X)$ is the lattice of opens of a topological space, joins are unions and finite meets are intersections.

Locales as generalised spaces

Locale theory abstracts away from topological spaces. The frame $\mathcal{O}(X)$ does not need to arise as a family of sets under union and intersection. Indeed there need not be any underlying set of points.

The generality can be motivated by considering frames as theories in a natural logic of observable properties, see Vickers (1989).

Many notions from topology transfer to locales, X, Y.

A (continuous) map $f\colon X \to Y$ is given by a function $f^{-1}\colon \mathcal{O}(Y) \to \mathcal{O}(X)$ that preserves arbitrary joins and finite meets.

A map $f\colon X \to Y$ is said to be an embedding if the function $f^{-1}\colon \mathcal{O}(Y) \to \mathcal{O}(X)$ is surjective.

The embeddings determine the notion of sublocale.

The two postulates interpreted for locales

Recall the two postulates.

(R1) For every open $U \subseteq \mathbf{2}^\omega$, if $\lambda(U) = 1$ then $\mathbf{R} \subseteq U$.

(R2) For every open $U \subseteq \mathbf{2}^\omega$, if $\mathbf{R} \subseteq U$ then $\lambda(U) = 1$.

$\mathbf{2}^\omega$ is a topological space, hence a locale. The open subsets $U \subseteq \mathbf{2}^\omega$ are the elements of $\mathcal{O}(\mathbf{2}^\omega)$, they determine the open sublocales of $\mathbf{2}^\omega$.

By interpreting $\subseteq$ as the sublocale relationship, the postulates make sense verbatim for locales, but the inconsistency vanishes.

Theorem *There exists a unique sublocale* $\mathbf{R} \subseteq \mathbf{2}^\omega$ *satisfying (R1) and (R2).*

Concrete description of $\mathbf{R}$

For $U, V \subseteq \mathbf{2}^\omega$ open, define:

$$U \approx V \;:\Leftrightarrow\; \lambda(U) = \lambda(U \cap V) = \lambda(V)$$

If $U \approx V$ then intuitively they represent the same observation on random sequences.

Define $\mathcal{O}(\mathbf{R}) \;:=\; \mathcal{O}(\mathbf{2}^\omega)/\approx$.

(Cf. the measure (σ-)algebra of $\mathbf{2}^\omega$, equivalently of $[0,1]$.)

Proposition *The above defines* $\mathbf{R}$ *as a sublocale of* $\mathbf{2}^\omega$*. This is the sublocale characterised by the previous theorem.*

Outer measure

We postulated that random sequences satisfy all measure 1 observable properties. What about other measure 1 properties?

The outer measure of a sublocale $Y \subseteq \mathbf{2}^\omega$ is defined by:

$$\lambda^*(Y) \;:=\; \inf\{\lambda(U) \mid U \in \mathcal{O}(\mathbf{2}^\omega) \text{ and } Y \subseteq U\}$$

(R2) asserts that $\mathbf{R}$ has outer measure 1.

Proposition $\mathbf{R}$ *is characterised as the smallest sublocale of* $\mathbf{2}^\omega$ *of outer measure* 1*.*

In particular, random sequences satisfy all measure 1 properties.

In contrast, in algorithmic randomness, the assumption that a sequence satisfies all measure 1 effective open properties (i.e., that it is Kurtz random) does not even imply the law of large numbers.

What are the random sequences in $\mathbf{R}$?

The locale $\mathbf{R}$ has no points, that is there are no maps from the one point (i.e., terminal) locale to $\mathbf{R}$.

Thus, according to this theory, there is no such thing as a completed infinite random sequence.

In particular, 0^ω is not random. Also, the bit-sequence of Chaitin's Ω is not random.

Nonethless $\mathbf{R}$ is a nontrivial space of random sequences.

(More generally, $\emptyset$ is the only compact sublocale of $\mathbf{R}$.)

What are the maps from $\mathbf{R}$ to $\mathbf{R}$?

Intuitively, these correspond to continuous transducers on random sequences.

Theorem *The maps* $\mathbf{R} \to \mathbf{R}$ *are in one-to-one correspondence with the continuous nonsingular maps from* $\mathbf{2}^\omega$ *to* $\mathbf{2}^\omega$ *modulo almost everywhere equivalence.*

A continuous nonsingular map from $\mathbf{2}^\omega$ to $\mathbf{2}^\omega$ is a continuous function from a measure 1 subspace $D \subseteq \mathbf{2}^\omega$ to $\mathbf{2}^\omega$, satisfying: for every null $Z \subseteq \mathbf{2}^\omega$, it holds that $f^{-1}(Z)$ is null.

More generally

If we generalise λ to other probability "measures" (technically, probability valuations), and $\mathbf{2}^\omega$ to other locales, we still obtain canonical "random" sublocales.

A (continuous) probability valuation on a locale X is a (necessarily monotone) function $\mu\colon \mathcal{O}(X) \to [0,1]$ satisfying:

$$\mu(\emptyset) = 0 \qquad \mu(U \cup V) = \mu(U) + \mu(V) - \mu(U \cap V)$$

$$\mu(X) = 1 \qquad \mu(\bigcup\nolimits^{\uparrow}_{U \in \mathcal{D}} U) = \sup_{U \in \mathcal{D}} \mu(U) \quad (\mathcal{D} \subseteq \mathcal{O}(X) \text{ directed})$$

Given (X, μ) define:

$$U \approx_\mu V \;:\Leftrightarrow\; \mu(U \cup V) = \mu(U \cap V)$$

$$\mathcal{O}(\mathbf{R}(\mu)) \;:=\; \mathcal{O}(X)/\approx_\mu$$

Proposition $\mathbf{R}(\mu)$ *is a sublocale of* X*.*

Define the outer value of a sublocale $Y \subseteq X$ by:

$$\mu^*(Y) := \inf\{\mu(U) \mid U \in \mathcal{O}(X) \text{ and } Y \subseteq U\}$$

Theorem *If the locale* X *is regular then:*

1. $\mathbf{R}(\mu)$ *is the meet of all open sublocales* $U \subseteq X$ *with* $\mu(U) = 1$
2. $\mathbf{R}(\mu)$ *is the meet of all sublocales* $Y \subseteq X$ *with* $\mu^*(Y) = 1$
3. $\mu^*(\mathbf{R}(\mu)) = 1$

Obviously, μ is also a probability valuation on $\mathbf{R}(\mu)$. Then $(\mathbf{R}(\mu), \mu)$ is random in the sense that:

$$[U] \subset [V] \in \mathcal{O}(\mathbf{R}(\mu)) \text{ (proper inclusion)} \Rightarrow \mu([U]) < \mu([V])$$

When does one measure subsume another?

For $\mathbf{2}^\omega$, valuations are in one-one correspondence with Borel measures.

Proposition *Given two probability measures μ, ν on $\mathbf{2}^\omega$, the following are equivalent:*

- $\mathbf{R}(\mu) \subseteq \mathbf{R}(\nu)$ *(as sublocales of $\mathbf{2}^\omega$)*
- $\mu \ll \nu$ *(i.e., μ is absolutely continuous relative to ν)*

Recall, μ is absolutely continuous relative to ν if every ν-null Borel set is also μ-null.

van Lambalgen's brainteaser revisited

Suppose we toss a sequence of biased coins, where the i-th coin has probability $\frac{1}{2} + \delta_i$ of coming up 0 and $\frac{1}{2} - \delta_i$ of coming up 1. Under what conditions on (δ_i) are we guaranteed to obtain a fair random sequence?

We write π for the product measure $\prod_i (\frac{1}{2} + \delta_i, \frac{1}{2} - \delta_i)$ on $\mathbf{2}^\omega$.

The question asks: When does it hold that $\mathbf{R}(\pi) \subseteq \mathbf{R}$?

Solution:

$$\mathbf{R}(\pi) \subseteq \mathbf{R} \iff \pi \ll \lambda \qquad \text{(by previous proposition)}$$

$$\iff \sum_{i=0}^{\infty} {\delta_i}^2 < \infty \qquad \text{(by Kakutani's Theorem)}$$

Another characterisation of $\mathbf{R}$

The locale of random sequences, $\mathbf{R}$, is:

- countably based
- zero dimensional
- and has no points

Moreover, the valuation λ on $\mathbf{R}$ is random.

Canonicity Theorem *If X is a countably-based zero-dimensional locale with no points and X carries a random probability valuation μ then X and $\mathbf{R}$ are homeomorphic via valuation preserving maps.*

(Cf. the measure algebra isomorphism theorem of measure theory.)

Constructive version?

Should be possible to develop the theory within Bishop Constructive Mathematics, and hence in any brand of constructivism.

In contrast, the usual theories of algorithmic randomness are not compatible with (formal) Church's Thesis.

In a constructive version, an existence proof for a map $\mathbf{R} \to \mathbf{R}$ should give rise to an effective transducer of random sequences.

Given a non-atomic effective probability valuation μ on $\mathbf{2}^\omega$, a constructive proof of the canonicity theorem should give mutually inverse transducers $\mathbf{R}(\mu) \to \mathbf{R}$ and $\mathbf{R} \to \mathbf{R}(\mu)$; the former amounting to an optimal data compression, and the latter its decoding. Thus a constructive canonicity theorem should relate to Shannon's noiseless coding theorem from information theory.

Random space?

If space S is continuous, how does this account for the existence of discrete observables $S \to \{0, 1\}$?

If space is discrete, how does this account for its geometry, e.g., straight lines, geodesics, etc.?

Perhaps space is random ...

A motivating idea for this is that any attempt to zoom in indefinitely on the location of a particle must eventually, at some level of granularity, settle down as a sequence of stochastic events.

Random space!

Let $\mathbf{R}(\mathbb{E}_n)$ be the smallest thick sublocale of $\mathbb{E}_n$ (Euclidean n-space).

Equivalently

$$\mathbf{R}(\mathbb{E}_n) = \overbrace{\mathbf{R}(\mathbb{E})\# \ldots \#\mathbf{R}(\mathbb{E})}^{n}$$

where $\#$ is the independent product of random locales.

The space $\mathbf{R}(\mathbb{E}_n)$ is zero dimensional, so have many nontrivial discrete observables $\mathbf{R}(\mathbb{E}_n) \to \{0, 1\}$.

$\mathbf{R}(\mathbb{E}_n)$ also possesses a nontrivial geometry. For example, it has random straight lines through independent points:

$$(x, y, \lambda) \mapsto \lambda x + (1 - \lambda)y \colon \ \mathbf{R}(\mathbb{E}_n)\#\mathbf{R}(\mathbb{E}_n)\#\mathbf{R}(\mathbb{E}) \ \to \ \mathbf{R}(\mathbb{E}_n)$$

Other directions

- Ideally, the development presented here should (once constructivised) be part of a general constructive account of probability and measure theory over locales. Aspects of such a theory have been initiated by, amongst others, Coquand, Palmgren, Spitters and Vickers. Much remains to be done.

- In recent talks, Dana Scott has identified the frame $\mathcal{O}(\mathbf{R})$ as a subframe of the measure algebra, and used this to give a sheaf model of a (classical) modal set theory (MZF) in which every proposition has a probability. This opens the possibility of performing modal set-theoretic reasoning about randomness.